

A near-optimal maintenance policy for automated DR devices

Carlos Abad and Garud Iyengar

Abstract—Demand side participation is now widely recognized as being extremely critical for satisfying the growing electricity demand in the US. The primary mechanism for demand management in the US is demand response (DR) programs that attempt to reduce or shift demand by giving incentives to participating customers via price discounts or rebate payments. Utilities that offer DR programs rely on automated DR devices (ADRs) to automate the response to DR signals. The ADRs are faulty; but the working state of the ADR is not directly observable – one can, however, attempt to infer it from the power consumption during DR events. The utility loses revenue when a malfunctioning ADR does not respond to a DR signal; however, sending a maintenance crew to check and reset the ADR also incurs costs. In this paper, we show that the problem of maintaining a pool of ADRs using a limited number of maintenance crews can be formulated as a restless bandit problem, and that one can compute a near-optimal policy for this problem using Whittle indices. We show that the Whittle indices can be efficiently computed using a variational Bayes procedure even when the load-shed magnitude is noisy and when there is a random mismatch between the clocks at the utility and at the meter. The results of our numerical experiments suggest that the Whittle-index based approximate policy is within 3.95% of the optimal solution for all reasonably low values of the signal-to-noise ratio in the meter readings.

NOMENCLATURE

Indices

- t Index of DR events
- k Index of discrete points in belief space $[0, 1]$
- i Index of meter readings during DR event
- j Index of samples of meter reading vectors

Constants and parameters

- λ Expected dollar savings for utility when ADR is working
- c Cost incurred by utility for sending a repair crew
- θ Customer compensation for participating in a DR event
- p Prior probability of ADR failure
- \mathcal{S} ADR state space
- γ_0 Non-operational ADR state
- γ_1 Operational ADR state
- \mathcal{A} Set of actions available to utility before a DR event
- α_0 Do-nothing action
- α_1 Send-crew-to-reset-ADR action
- \mathcal{X} Set of possible meter reading vectors during DR event
- β Discount factor between DR events
- n Discretization size of belief space $[0, 1]$
- D Number of ADRs managed by utility

- M Number of repair crews overseeing ADRs
- m Number of meter readings during DR event
- N Number of samples of meter reading vectors
- \mathbf{y} Vector of estimated normal power consumption
- σ Standard deviation of load estimation residuals
- ν_0 Expected load-shed during a DR event
- η_0 Load-shed precision (inverse of variance)
- d Maximum absolute meter-utility clock mismatch

Variables

- s State of ADR before DR event
- a Action taken by utility before DR event
- \mathbf{x} Vector of meter readings during DR event
- b Belief probability that ADR is operational
- $\tilde{\epsilon}$ Load estimation residual
- r Actual load-shed during a DR event
- δ Actual mismatch between meter and utility clocks
- ω Low-discrepancy sample in unit hypercube
- \mathbf{z} Vector of IID standard normal samples

Functions

- R_{as} Utility profit when taking action a in ADR state s
- $P_{ss'}(a)$ ADR transition probability
- $Q_{sa}(\mathbf{x})$ Meter reading observation conditional probability
- $r_a(b)$ Expected profit under action a in belief state b
- $\Gamma_{ax}(b)$ Belief probability transition map
- $W_{ab}(\mathbf{x})$ Up-to-date meter reading observation probability
- $V(b)$ POMDP value function
- $\Phi(\cdot)$ Standard normal cumulative distribution function
- $\varphi(\cdot)$ Standard normal density function

I. INTRODUCTION

Demand Response (DR) refers to a set of activities where end-use customers change or shift their normal electricity consumption patterns in response to changes in the price of electricity or other incentive payments [1]. DR has been extensively deployed over the past several years to improve electric grid reliability and market efficiency [2], and it is expected to play an even more prominent role with the planned integration of intermittent renewable generation. There are three levels of DR automation: manual DR —where each equipment controller is manually turned off; semi-automated DR —where an individual triggers a preprogrammed DR strategy via a centralized control system; and fully automated DR —where the DR strategy is initiated by an automated DR device (ADR) on receipt of an external communications signal [3]. The utilities offering DR programs prefer to install ADRs since the fully automated system reduces the operating costs of DR programs by increasing DR resource reliability and reducing the amount of effort required from end-use customers [4].

Supported in part by DOE grant de-ar0000235, ONR grant N000140310514, and NSF grant DMS-1016571

Both authors are with the Industrial Engineering and Operations Research Department, Columbia University, New York, NY 10027, USA
ca2446@columbia.edu, garud@ieor.columbia.edu

ADRs are faulty [5],[6], and have to be periodically inspected and reset by sending a maintenance crew. Although newer ADRs are equipped with a two-way communication that allow the utility to remotely observe the ADR state, the vast majority of the deployed ADRs use one-way communication technology and, therefore, their state is not directly observable. According to some estimates, failure to identify non-functioning ADRs can reduce the effectiveness of DR programs by approximately 20-30% and lead to lost revenues of the order of \$ 1.7M for a utility with 1M customers and 10% participation rate [7]. Consequently, identifying malfunctioning ADRs is of immense economic value for the utilities. The current practice adopted by utilities is to regularly send a maintenance team to inspect and possibly repair the ADRs. In this paper we propose a method that is able to infer the ADR state from meter readings, and use the estimates to optimally schedule the ADR maintenance. We show that our proposed method clearly outperforms any regular maintenance schedule, without the need of any new investment in additional hardware. The method we propose can also be used by utilities that have invested in two-way communication ADRs to verify whether the reported ADR state was accurate.

As a first step towards solving the ADR repair scheduling problem, we formulate the maintenance problem for a single ADR using only noisy meter readings. We assume that an ADR can be in either in a *functional* or a *non-functional* state. We assume that, over a given time, the ADR state transitions from a functional to a non-functional state with a known probability. Thus, state transitions form a Markov chain. The utility decides whether to *send* a crew to reset the device or *do nothing*. This decision would be simple if the true ADR state was directly observable without any errors. In most currently deployed ADRs, the state is not observable; it can only be inferred from the customer's noisy electricity consumption. Thus, the ADR maintenance problem can be modeled as a Markov decision process (MDP) with partially observable states (POMDP). The POMDP framework incorporates the uncertainty associated with any estimation process into a "belief" probability that the ADR is functioning, and provides a methodology for updating this belief as more information becomes available, i.e. more meter readings are recorded during new DR events. POMDPs are, in general, very hard optimization problems [8], [9]. However, it is often possible to compute the optimal policy for POMDPs with small state, action and observation spaces, or additional structure. We show that the optimal policy in our problem is a single threshold policy where it is optimal to send a maintenance crew whenever the belief probability drops below the threshold value. We show that the optimal threshold can be approximated to any degree of accuracy by solving a single linear program (LP).

Our model for a single ADR maintenance problem falls in the class of random failure models. Currently existing alternatives to our approach are empirical predictive maintenance routines such as the Reliability-Centered Maintenance, and deterioration failure models; see [10] for details. These methodologies are typically used for managing traditional electric assets such as generators and transformers. To the best of our knowledge, there is no previous work on probabilistic

models for the maintenance of a single ADR, let alone the maintenance scheduling of multiple ADRs being overseen by a small number of repair teams.

We formulate the problem of maintaining a pool of ADRs using a limited number of maintenance crews as a restless bandit (RB) problem [11]. The RB problem is a generalization of the multiarmed bandit (MAB) problem [12]. In the MAB problem, the decision maker chooses a set of "bandits" to activate based on the current state information, and the state of the chosen bandits evolves according to a known distribution; however, the states of all the inactive bandits remain fixed. Gittins [12] constructed a set of index functions that map the state of a bandit to a real number, and showed that the optimal solution for the MAB problem is to select the bandits in the order determined by the index function. In the RB setting, the states of the inactive bandits also evolve. In the ADR maintenance setting, an active "bandit" corresponds to an ADR that is being inspected and possibly repaired, and an inactive "bandit" corresponds to an ADR that is not being inspected; clearly, the state of inactive ADRs continue to evolve according to its failure distribution. Computing optimal policies for the RB problem is hard [13], but a generalization of Gittins' indices can be used for constructing approximately optimal and very efficiently implementable policies for a number of applications. Whittle [11] established conditions under which one can define generalized Gittins' indices for the RB problem. We will refer to such RB problems as Whittle-indexable. Glazebrook et al. [14] have established that machine maintenance models that are either monotone or are breakdown/deterioration models, are Whittle-indexable. The ADR repair scheduling problem is neither monotone nor a breakdown/deterioration model; therefore, the results in [14] do not extend to this model. We establish that the ADR repair scheduling problem is Whittle indexable, and show that the Whittle indices can be computed by solving a sequence of single ADR maintenance problems.

Finally, we conduct an extensive numerical study where we explore several additional practical issues such as the impact on performance when the utility and meter clock are not synchronous, and the impact of uncertain DR load shed. Our proposed variational Bayes procedure to handle these issues is of independent interest. The results of our numerical experiments suggest that, for reasonable values of the signal-to-noise ratio in the meter readings, the Whittle-index policy is within 3.95% of the optimal policy.

II. ADR MAINTENANCE PROBLEM

In this section we investigate the maintenance problem for a single ADR. The solution to this problem will be used to solve a relaxation of the multiple ADRs repair scheduling problem in Section III.

For $t \geq 1$, let $s_t \in \mathcal{S}$ denote the ADR state just prior to the t -th DR event, let $a_t \in \mathcal{A}$ denote the action taken just prior to the t -th DR event, and let \mathbf{x}_t denote the vector of meter readings recorded during the t -th DR event. During DR event t , the utility receives a profit $R_{a_t s_t}$. The profit matrix \mathbf{R} is

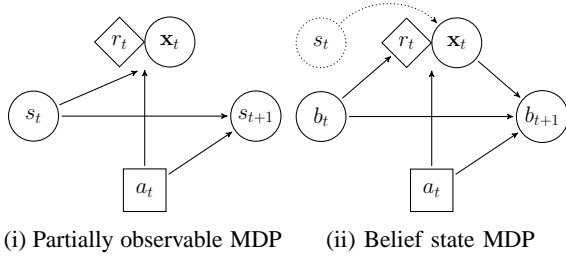


Fig. 1: ADR maintenance problem causal relationships

given by

$$\mathbf{R} = [R_{as}]_{a \in \mathcal{A}, s \in \mathcal{S}} = \begin{matrix} & \gamma_0 & \gamma_1 \\ \alpha_0 & \begin{pmatrix} -\theta & \lambda - \theta \\ \lambda - \theta - c & \lambda - \theta - c \end{pmatrix} \\ \alpha_1 & \end{matrix} \quad (1)$$

For ease of notation, and w.l.o.g., we will assume that $\theta = 0$. We assume that an ADR that was functioning during the t -th event, i.e. $s_t = \gamma_1$, fails before the $(t+1)$ -th DR event, i.e. $s_{t+1} = \gamma_0$, with probability p . Thus, under the do-nothing action α_0 , the state transition matrix

$$\mathbf{P}(\alpha_0) = [P_{ss'}(\alpha_0)]_{ss' \in \mathcal{S}} = \begin{matrix} & \gamma_0 & \gamma_1 \\ \gamma_0 & \begin{pmatrix} 1 & 0 \\ p & 1-p \end{pmatrix} \\ \gamma_1 & \end{matrix}. \quad (2)$$

Since the action α_1 resets the ADR state to γ_1 just prior to the t -th DR event, it follows that the state transition matrix

$$\mathbf{P}(\alpha_1) = [P_{ss'}(\alpha_1)]_{ss' \in \mathcal{S}} = \begin{matrix} & \gamma_0 & \gamma_1 \\ \gamma_0 & \begin{pmatrix} p & 1-p \\ p & 1-p \end{pmatrix} \\ \gamma_1 & \end{matrix}. \quad (3)$$

We assume that the true ADR state can only be determined by sending a maintenance crew to inspect it. However, the utility can use the vector \mathbf{x} of metered power consumption to infer the state of the ADR. Let

$$Q_{sa}(\mathbf{x}) = \mathbb{P}(\mathbf{x}_t = \mathbf{x} \mid a_t = a, s_t = s) \quad (4)$$

denote the conditional probability of observing the vector of meter readings $\mathbf{x}_t = \mathbf{x}$ during the t -th DR event when action $a_t = a$, and the ADR state $s_t = s$. The tuple $(\mathcal{S}, \mathcal{A}, \mathcal{X}, \mathbf{P}, \mathbf{R}, \mathbf{Q})$ is a partially observable Markov decision process (POMDP) that completely describes the ADR maintenance problem. Note that we deviate from the standard definition of POMDP (see, e.g. [15]) wherein the observation \mathbf{x}_t is a function of the current state s_t and the *previous* action a_{t-1} . We do so in order to keep the dynamics in the problem more transparent. Figure 1 describes the causal relationships between states, actions, rewards, and observations in the ADR maintenance problem.

It is well-known [16] that a POMDP is equivalent to a belief state MDP. Let $h_t = \{a_1, \mathbf{x}_1, a_2, \mathbf{x}_2, \dots, a_{t-1}, \mathbf{x}_{t-1}\}$ denote the observed history before the t -th DR event. Then,

$$b_t = \mathbb{P}(s_t = \gamma_1 \mid h_t) \in \mathcal{B} = [0, 1] \quad (5)$$

denotes the belief probability that the ADR is operational during the t -th DR event. The expected profit $\mathbf{r}(b)$ in belief state $b \in \mathcal{B}$ is given by

$$\mathbf{r}(b) = [r_{\alpha_0}(b), r_{\alpha_1}(b)]^\top = \mathbf{R} [1 - b, b]^\top = [\lambda b, \lambda - c]^\top. \quad (6)$$

Note that the updated belief state b_{t+1} after observing the meter readings \mathbf{x}_t is $b_{t+1} = \mathbb{P}(s_{t+1} = \gamma_1 \mid h_t, a_t = a, \mathbf{x}_t = \mathbf{x})$. Using Bayes' rule and the causality structure in Figure 1, one can show that $b_{t+1} = \Gamma_{a\mathbf{x}}(b)$, where the map

$$\Gamma_{a\mathbf{x}}(b) = \frac{P_{\gamma_1\gamma_1}(a)Q_{\gamma_1a}(\mathbf{x})b + P_{\gamma_0\gamma_1}(a)Q_{\gamma_0a}(\mathbf{x})(1-b)}{W_{ab}(\mathbf{x})}, \quad (7)$$

and $W_{ab}(\mathbf{x}) = \mathbb{P}(\mathbf{x}_t = \mathbf{x} \mid h_t, a_t = a) = Q_{\gamma_1a}(\mathbf{x})b + Q_{\gamma_0a}(\mathbf{x})(1-b)$. The belief transition (7) is standard for POMDPs; see, e.g. [17] for details. The tuple $(\mathcal{B}, \mathcal{A}, \Gamma, \mathbf{r})$ is the belief-state MDP that represents the ADR maintenance problem.

A. Optimal policy

Our goal is to compute a policy $\pi^* : \mathcal{B} \rightarrow \mathcal{A}$ that maximizes the total expected discounted profit

$$V^\pi(b_1) = \sum_{t=1}^{\infty} \beta^{t-1} r_{a_t^\pi}(b_t), \quad (8)$$

where $\beta \in (0, 1)$ is a discount factor and a_t^π denotes the action taken by policy π in belief state b_t prior to the t -th DR event. It is well known [18] that there exists a stationary optimal policy $\pi^* \in \arg\max_{\pi} V^\pi(b_1)$, and that the associated optimal value function $V^{\pi^*}(b)$, satisfies the Bellman equation:

$$V(b) = \max_{a \in \mathcal{A}} \left\{ r_a(b) + \beta \mathbb{E}_{\mathbf{x}} [V(\Gamma_{a\mathbf{x}}(b))] \right\}, \quad (9)$$

where $\mathbb{E}_{\mathbf{x}} [V(\Gamma_{a\mathbf{x}}(b))] = \int_{\mathbf{x} \in \mathcal{X}} V(\Gamma_{a\mathbf{x}}(b)) W_{ab}(\mathbf{x}) d\mathbf{x}$ denotes the expected future profit. From (9) it follows that action α_1 is optimal for belief state b if the function

$$f(b) := r_{\alpha_0}(b) - r_{\alpha_1}(b) + \beta \mathbb{E}_{\mathbf{x}} [V(\Gamma_{\alpha_0\mathbf{x}}(b))] - \beta \mathbb{E}_{\mathbf{x}} [V(\Gamma_{\alpha_1\mathbf{x}}(b))] \leq 0. \quad (10)$$

We use the following result to show that the optimal policy is a single threshold policy.

Lemma 1. *The function f is convex in $[0, 1]$.*

Proof: Note that $r_{\alpha_0}(b) = \lambda b$ is an affine function of b and $r_{\alpha_1}(b) = \lambda - c$ is independent of b . It follows that $r_{\alpha_0}(b) - r_{\alpha_1}(b)$ is an affine and, therefore, convex function of b . Also, since $\Gamma_{\alpha_1\mathbf{x}}(b) = 1 - p$, the last term $\mathbb{E}_{\mathbf{x}} [V(\Gamma_{\alpha_1\mathbf{x}}(b))]$ in $f(b)$ is, in fact, independent of b .

The value function V of an infinite horizon MDP is convex on $[0, 1]$ [18]. Then, the perspective function $g(v, u) = uV(v/u)$ is convex on $\{(v, u) : v/u \in [0, 1], v \in \mathbb{R}_+\}$ [19]. Since $W_{\alpha_0b}(\mathbf{x}) = Q_{\gamma_1\alpha_0}(\mathbf{x})b + Q_{\gamma_0\alpha_0}(\mathbf{x})(1-b)$ and

$$\Gamma_{\alpha_0\mathbf{x}}(b) = \frac{(1-p)Q_{\gamma_1\alpha_0}(\mathbf{x})b}{W_{\alpha_0b}(\mathbf{x})}, \quad (11)$$

it follows that

$$g_{\mathbf{x}}(b) := V(\Gamma_{\alpha_0 \mathbf{x}}(b))W_{\alpha_0 \mathbf{x}}(b) \quad (12)$$

$$= g\left((1-p)Q_{\gamma_1 \alpha_0}(\mathbf{x})b, Q_{\gamma_1 \alpha_0}(\mathbf{x})b + Q_{\gamma_0 \alpha_0}(\mathbf{x})(1-b)\right)$$

is a convex function of b for each $\mathbf{x} \in \mathcal{X}$. Hence, $\mathbb{E}_{\mathbf{x}}[V(\Gamma_{\alpha_0 \mathbf{x}}(b))] = \int_{\mathbf{x} \in \mathcal{X}} g_{\mathbf{x}}(b)d\mathbf{x}$ is a convex function of b . ■

Theorem 1. *The optimal policy π^* for the ADR maintenance problem is a threshold policy, i.e. there exists $b^* \in \mathbb{R}$ such that*

$$a_t^* = \begin{cases} \alpha_1 & \text{if } b_t \leq b^* \\ \alpha_0 & \text{if } b_t > b^*. \end{cases} \quad (13)$$

Proof: From Lemma 1, we have that $f(b)$ is convex. Consequently, the set $I = \{b : f(b) \leq 0\}$ of belief states for which action α_1 is optimal is a closed, possibly empty, interval. Suppose I is empty. Then the optimal policy is of the form (13) with $b^* < 0$.

Next, suppose I is non-empty. To establish the result, it suffices to show that $0 \in I$. Suppose not, i.e. α_0 is strictly optimal for $b = 0$. Then,

$$V(0) = r_{\alpha_0}(0) + \beta \mathbb{E}_{\mathbf{x}}[V(\Gamma_{\alpha_0 \mathbf{x}}(0))] = 0 + \beta V(0), \quad (14)$$

where we use fact that $\Gamma_{\alpha_0 \mathbf{x}}(0) = 0$. It follows that $V(0) = 0$. Moreover,

$$0 = V(0) > r_{\alpha_1}(0) + \beta \mathbb{E}_{\mathbf{x}}[V(\Gamma_{\alpha_1 \mathbf{x}}(0))] = \lambda - c + \beta V(1-p). \quad (15)$$

Then, for any belief state b ,

$$r_{\alpha_1}(b) + \beta \mathbb{E}_{\mathbf{x}}[V(\Gamma_{\alpha_1 \mathbf{x}}(b))] = \lambda - c + \beta V(1-p) < 0. \quad (16)$$

Since the payoff from action α_0 is non-negative in any belief state, we have that $V(b) \geq 0$ for all $b \in \mathcal{B}$. It follows that $V(b) = r_{\alpha_0}(b) + \beta \mathbb{E}_{\mathbf{x}}[V(\Gamma_{\alpha_0 \mathbf{x}}(b))]$, and α_0 is the unique optimal action for all belief states. Thus, we have established that the interval I is empty; a contradiction. Hence, $0 \in I$ and the optimal policy is of the form (13) with $b^* \in [0, 1]$. ■

We conclude this section with the following corollary.

Corollary 1. *The level sets $\{b : f(b) \leq -\mu\}$ of the function f are all of the form $[0, \xi]$, where we use the convention that $[0, \xi] = \emptyset$ if $\xi < 0$.*

Proof: Consider an ADR maintenance problem where the reward associated with the do-nothing action α_0 is increased by an amount μ . Then the interval over which it is optimal to take action α_1 is given by $\{b : f(b) + \mu \leq 0\}$. From the proof of Theorem 1, it follows that the set $\{b : f(b) \leq -\mu\}$ is of the form $[0, \xi]$. ■

We use Corollary 1 in Section III to establish that the ADR repair scheduling problem is Whittle-indexable [14] and, therefore, that there exists a well-defined heuristic policy for approximately solving it.

B. Approximating the optimal threshold

In this section, we describe a numerical scheme to approximate the optimal threshold b^* . We discretize the space $\mathcal{B} = [0, 1]$ into $n + 1$ equally spaced points $\hat{\mathcal{B}} = \{k/n : k = 0, \dots, n\}$, and round up $b \in ((k-1)/n, k/n]$ to the point k/n . Formally, we consider the MDP with state space $\hat{\mathcal{B}}$, action space \mathcal{A} , reward function $\mathbf{r}|_{\hat{\mathcal{B}}}$ and state transition $\hat{\Gamma}_{\mathbf{ax}}(k) := \lceil \Gamma_{\mathbf{ax}}(k/n) \rceil$. Let $\hat{\mathbf{V}} = (\hat{V}(0), \dots, \hat{V}(k), \dots, \hat{V}(n))^T$ denote the value function vector for the $n + 1$ states $k/n \in \hat{\mathcal{B}}$. Then

$$\hat{V}(k) = \max \left\{ \begin{array}{l} r_{\alpha_0}(k/n) + \beta \mathbb{E}_{\mathbf{x}}[\hat{V}(\hat{\Gamma}_{\alpha_0 \mathbf{x}}(k))], \\ r_{\alpha_1}(k/n) + \beta \mathbb{E}_{\mathbf{x}}[\hat{V}(\hat{\Gamma}_{\alpha_1 \mathbf{x}}(k))] \end{array} \right\}. \quad (17)$$

It is well-known [20] that the vector $\hat{\mathbf{V}}$ can be computed by solving the LP:

$$\begin{aligned} \min_{\hat{\mathbf{V}}} \quad & \sum_{k=0}^n V(k) \\ \text{s.t.} \quad & V(k) \geq r_{\alpha_0}(k/n) + \beta \mathbb{E}_{\mathbf{x}}[V(\hat{\Gamma}_{\alpha_0 \mathbf{x}}(k))], \quad \forall k, \\ & V(k) \geq r_{\alpha_1}(k/n) + \beta V(\lceil (1-p)n \rceil), \quad \forall k, \end{aligned} \quad (18)$$

where we use the fact that $\hat{\Gamma}_{\alpha_1 \mathbf{x}}(k) = \lceil (1-p)n \rceil$. LP (18) consists of n variables and $2n$ constraints and, therefore, can be solved very fast, provided the conditional expectations $\mathbb{E}_{\mathbf{x}}[V(\hat{\Gamma}_{\alpha_0 \mathbf{x}}(k))]$ can be computed efficiently. In Section IV, we show how to efficiently approximate the conditional expectations in (18) using low discrepancy sequences. Given the approximate value function $\hat{\mathbf{V}}$, we approximate the threshold $b^* \approx \frac{1}{n} \max\{k : \hat{V}(k) = r_{\alpha_1}(k/n) + \beta \hat{V}(\lceil (1-p)n \rceil)\}$.

III. ADR REPAIR SCHEDULING PROBLEM

Suppose there are M maintenance crews available to maintain $D(\gg M)$ ADRs. Before each DR event t , the utility must decide which (if any) ADRs to inspect and reset. We formulate the ADR repair scheduling problem as a restless bandit (RB) problem [11] where each of the ADR devices is a bandit. Following the notation of bandit problems, we will call an ADR active if it is being inspected, and inactive otherwise.

For RB problems, Whittle [11] proposed a possibly suboptimal policy based on the Lagrangian relaxation for the dynamic program. The Lagrangian relaxation decouples the RB problem into a collection of so called subsidy- μ problems. For each bandit, the subsidy- μ problem is an ADR maintenance problem in which the reward under the do-nothing action α_0 is increased by an amount $\mu \geq 0$.

Let $\mathcal{B}_0(\mu)$ denote the set of states for which the do-nothing action α_0 is optimal for a subsidy level μ . An RB problem is said to be *indexable* if $\mathcal{B}_0(\mu) \subseteq \mathcal{B}_0(\zeta)$ whenever $\mu \leq \zeta$. For indexable RB problems, the Whittle index μ_{κ} of bandit κ in state b_{κ} is defined as the minimum subsidy $\mu \geq 0$ which makes the passive action α_0 optimal at b_{κ} . Let $\mathcal{K} = \{\kappa_{(\ell)} : \ell = 1, \dots, K\}$ denote the indices of bandits with strictly positive Whittle index arranged in decreasing order of the index, i.e. $\mu_{\kappa_{(\ell)}} \geq \mu_{\kappa_{(\iota)}}$ whenever $\ell \leq \iota$. The Whittle-index policy specifies that the set of bandits that are activated at time t is given by $\mathcal{K}_1 = \{\kappa_{(\ell)} : 1 \leq \ell \leq \min\{K, M\}\}$, i.e. at most

M bandits with the largest strictly positive Whittle index are activated. The next theorem ensures that the Whittle indices are well defined for our problem.

Theorem 2. *The ADR repair scheduling problem is indexable.*

Proof: The Bellman equation for the subsidy- μ problem corresponding to the ADR scheduling problem is given by

$$V_\mu(b) = \max_{a \in \mathcal{A}} \{r_a(b, \mu) + \beta \mathbb{E}_{\mathbf{x}}[V_\mu(\Gamma_{a\mathbf{x}}(b))]\}, \quad (19)$$

where $\mathbf{r}(b, \mu) = \mathbf{r}(b) + (\mu, 0)^\top$. The set of states for which α_0 is optimal is of the form $\mathcal{B}_0(\mu) = \{b \in [0, 1] : b > b^*(\mu)\}$ where $b^*(\mu)$ denotes the optimal threshold corresponding to the subsidy- μ problem (19). Hence, in order to establish that the ADR repair scheduling problem is indexable, we need to show that the threshold $b^*(\mu)$ is non-increasing in μ . From Corollary 1, it follows that $[0, b^*(\mu)] = \{b : f(b) \leq -\mu\}$. Thus, $b^*(\mu)$ is non-increasing in μ . ■

Since the Whittle index $\mu^*(b) = \inf\{\mu \geq 0 : b \in \mathcal{B}_0(\mu)\} = \inf\{\mu \geq 0 : b > b^*(\mu)\}$, it is clear that $\bar{\mu} = \inf\{\mu \geq 0 : b^*(\mu) < 0\}$ is an upper bound for $\mu^*(b)$. Therefore, the Whittle index for any $b \in \mathcal{B}$ can be computed to within an accuracy ϵ using a binary search by solving at most $\mathcal{O}(\log_2(\frac{\bar{\mu}}{\epsilon}))$ LPs of the form (18). Thus, in practice, one needs to solve at most $\mathcal{O}(n \log_2(\frac{\bar{\mu}}{\epsilon}))$ LPs. On the other hand, we can also compute Whittle indices to within ϵ accuracy by finding the optimal threshold $b^*(\mu)$ for all $\mu \in \mathcal{M} = \{k\epsilon : 0 \leq k \leq \lceil \frac{\bar{\mu}}{\epsilon} \rceil\}$. Thus, the overall complexity of computing the Whittle indices is at most $\mathcal{O}(\min\{n \log_2(\frac{\bar{\mu}}{\epsilon}), \lceil \frac{\bar{\mu}}{\epsilon} \rceil\})$. In our numerical experiments, we calculated a bound on $\bar{\mu}$ using a doubling strategy, and used the binary search approach to compute the Whittle indices.

IV. NUMERICAL IMPLEMENTATION

Recall that as long as we can compute the conditional expectation $\mathbb{E}_{\mathbf{x}}[V(\hat{\Gamma}_{\alpha_0 \mathbf{x}}(k))]$ efficiently, we can solve LP (18) and, consequently, implement the Whittle-index policy. In this section, we discuss how to efficiently compute the required conditional expectation in two situations that arise in the DR context: when the DR load-shed is random, and when the clocks at the utility and at the meter are not synchronized. We show that these issues can have a significant impact on the performance of the Whittle-index policy, and propose a variational Bayes procedure to address them. Finally, we evaluate the performance of our proposed policies with respect to periodic review policies, and with respect to policies that have full information of the state of the ADRs.

A. Deterministic load-shed and synchronized clocks

Let $\mathbf{x} = (x_1, \dots, x_m)$ denote the sampled power consumption over a DR event of length m periods. For $i = 1, \dots, m$, we assume that the power consumption x_i is given by

$$x_i = \begin{cases} y_i - r + \tilde{\varepsilon}_i & \text{if the ADR is operational} \\ y_i + \tilde{\varepsilon}_i & \text{otherwise,} \end{cases} \quad (20)$$

where y_i denotes the estimated power consumption on a non-DR day, r denotes the load-shed mandated by the utility, and the estimation residual $\tilde{\varepsilon}_i$ is assumed to be independent and

identically distributed (IID) according to a normal distribution with mean 0 and variance σ^2 . As will become clear below, we can work with any specification for the residuals ε_i as long as one is able to simulate samples from the distribution.

We approximate the conditional expectation $\mathbb{E}_{\mathbf{x}}[V(\hat{\Gamma}_{\alpha_0 \mathbf{x}}(k))]$ in the LP (18) by the finite weighted sum

$$\mathbb{E}_{\mathbf{x}}[V(\hat{\Gamma}_{\alpha_0 \mathbf{x}}(k))] \approx \frac{1}{N} \sum_{j=1}^N V(\hat{\Gamma}_{\alpha_0 \mathbf{x}^j}(k)), \quad (21)$$

where the samples $\{\mathbf{x}^j : j = 1, \dots, N\}$ are generated IID from the distribution $W_{\alpha_0, b}(\mathbf{x})$, $b = k/n$, i.e.

$$\mathbf{x}^j \sim \begin{cases} \mathcal{N}(\mathbf{y}, \sigma^2 \mathbf{I}) & \text{w.p. } 1 - b, \\ \mathcal{N}(\mathbf{y} - r\mathbf{1}, \sigma^2 \mathbf{I}) & \text{w.p. } b. \end{cases} \quad (22)$$

To efficiently sample from the multivariate Normal distributions, we use low-discrepancy sequences in the unit m -dimensional hypercube [21]. Let $\omega^1, \dots, \omega^N \in [0, 1]^m$ be the first N elements of a low-discrepancy sequence. We define $\mathbf{z}^j \in \mathbb{R}^m$ by setting $z_i^j = \Phi^{-1}(w_i^j)$. Then $\{\mathbf{z}^j : j = 1, \dots, N\}$ are IID samples from an m -dimensional standard Normal random variable [22]. We generate meter readings sample \mathbf{x}^j as follows:

$$\mathbf{x}^j = \begin{cases} \mathbf{y} + \sigma \mathbf{z}^j & \text{w.p. } 1 - b, \\ \mathbf{y} - r\mathbf{1} + \sigma \mathbf{z}^j & \text{w.p. } b. \end{cases} \quad (23)$$

Let $Q_0(\mathbf{x})$ (resp. $Q_1(\mathbf{x})$) denote the probability of observing a vector of meter readings \mathbf{x} under an functional (resp. non-functional) ADR. Then, $Q_{\gamma_0 \alpha_0}(\mathbf{x}) = Q_0(\mathbf{x})$, and $Q_{\gamma_0 \alpha_1}(\mathbf{x}) = Q_{\gamma_1 \alpha_0}(\mathbf{x}) = Q_{\gamma_1 \alpha_1}(\mathbf{x}) = Q_1(\mathbf{x})$, where

$$Q_0(\mathbf{x}) = \prod_{i=1}^m \varphi\left(\frac{x_i - y_i}{\sigma}\right), \quad Q_1(\mathbf{x}) = \prod_{i=1}^m \varphi\left(\frac{x_i - y_i + r}{\sigma}\right). \quad (24)$$

Given $Q_0(\mathbf{x}^j)$ and $Q_1(\mathbf{x}^j)$, we compute $\hat{\Gamma}_{\alpha_0, \mathbf{x}^j}(k) = \lceil \Gamma_{\alpha_0, \mathbf{x}^j}(b) \rceil$ using (7).

B. Random load-shed and unsynchronized clocks

Here, we assume that the distribution of the load-shed is known –however, the exact magnitude is unknown– and that the clock at the utility and in the meter can be mismatched up to d time units. We use the utility’s clock time as the reference time for all computations, and assume that meter readings are assigned to time instants using the meter clock; therefore, an observation assigned to the time instant i could, in fact, correspond to an actual instant in the set $\{i - d, \dots, i + d\}$. Suppose a DR event takes place during the period $\{d + 1, \dots, d + m\}$. Then,

$$\mathbf{x} \sim \begin{cases} \mathcal{N}(\mathbf{y} - r\mathbf{1}_\delta, \sigma^2 \mathbf{I}) & \text{for some } r, \delta \in \{-d, \dots, d\} \\ & \text{when the ADR is operational,} \\ \mathcal{N}(\mathbf{y}, \sigma^2 \mathbf{I}) & \text{when the ADR is not operational,} \end{cases} \quad (25)$$

where $\mathbf{1}_\delta \in \mathbb{R}^{m+2d}$ denotes a vector with the components $i \in I_\delta = \{1 + d + \delta, \dots, m + d + \delta\}$ equal to 1, and all other components equal to zero.

Let $\omega^1, \dots, \omega^N$ be the first N elements of a low-discrepancy sequence in the unit $(m + 2d)$ -dimensional hypercube and $z_i^j = \Phi^{-1}(w_i^j)$. Then, the j -th sample from the distribution $W_{\alpha_0, b}(\mathbf{x}^j)$ is given by

$$\mathbf{x}^j = \begin{cases} \mathbf{y} + \sigma \mathbf{z}^j & \text{w.p. } 1 - b, \\ \mathbf{y} - r \mathbf{1}_\delta + \sigma \mathbf{z}^j & \text{w.p. } b\rho(r, \delta), \end{cases} \quad (26)$$

where ρ denotes the joint probability distribution function of the load-shed and the clock mismatch. In this case, the observation probabilities are

$$Q_0(\mathbf{x}) = \prod_{i=1}^{m+2d} \varphi\left(\frac{x_i - y_i}{\sigma}\right), \quad (27)$$

$$Q_1(\mathbf{x}) = \mathbb{E}_{\mathbf{x}} \left[\prod_{i \in I_\delta} \varphi\left(\frac{x_i - y_i + r}{\sigma}\right) \prod_{i \notin I_\delta} \varphi\left(\frac{x_i - y_i}{\sigma}\right) \right],$$

where the expectation is with respect to the posterior distribution of the mismatch δ and the load-shed r . In the next section, we show how to use the variational Bayes procedure to approximate the posterior expectation. Given $Q_1(\mathbf{x}^j)$ and $Q_0(\mathbf{x}^j)$, we approximate $\mathbb{E}_{\mathbf{x}}[V(\hat{\Gamma}_{\alpha_0 \mathbf{x}}(k))]$ using (21) and (7).

C. Belief state update using variational Bayes

Consider the following hierarchical Bayesian model

$$\begin{aligned} \mathbf{x} &= \mathbf{y} - r \mathbf{1}_\delta + \varepsilon, & r &\sim \mathcal{N}(\nu_0, 1/\eta_0), \\ \delta &\sim \text{unif}\{-d, \dots, d\}, & \varepsilon &\sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}), \end{aligned} \quad (28)$$

where ν_0 is the expected load-shed during the DR event, and $\eta_0^{-1/2}$ is its standard deviation. The posterior distribution $\rho(r, \delta | \mathbf{x}, \mathcal{H})$ is proportional to $\mathbb{P}(\mathbf{x} | r, \delta, \sigma) \mathbb{P}(r | \nu_0, \eta_0) \mathbb{P}(\delta)$, where

$$\mathbb{P}(\mathbf{x} | \delta, r, \sigma) = \prod_{i \in I_\delta} \varphi\left(\frac{x_i - y_i + r}{\sigma}\right) \prod_{i \notin I_\delta} \varphi\left(\frac{x_i - y_i}{\sigma}\right). \quad (29)$$

Since this posterior distribution is neither in closed form, nor is it easy to sample from, we use the variational inference method [23] and approximate the posterior distribution by the product distribution $\Delta(\delta)g(r)$. It is easy to establish that the product distribution that minimizes the Kullback-Leibler distance from the joint posterior distribution is of the form

$$\Delta(\delta) \propto \exp \left\{ \frac{\nu}{\sigma^2} \mathbf{1}_\delta^\top (\mathbf{y} - \mathbf{x}) - \frac{\nu^2 + \eta^{-1}}{2\sigma^2} \|\mathbf{1}_\delta\|_2^2 \right\} \quad (30)$$

$$g(r) = \mathcal{N}(\nu, 1/\eta), \quad (31)$$

where $\eta = \eta_0 + \frac{1}{\sigma^2} \sum_{\delta=-d}^d \Delta(\delta) \|\mathbf{1}_\delta\|_2^2$ and $\nu = \eta^{-1} \left(\nu_0 \eta_0 + \frac{1}{\sigma^2} \sum_{\delta=-d}^d \Delta(\delta) \mathbf{1}_\delta^\top (\mathbf{y} - \mathbf{x}) \right)$. Note the circular dependence of the parameters in the posterior distributions. We use an iterative procedure that alternates between $\Delta(\delta)$, and the precision η and mean ν of the normal distribution for r . We terminate the procedure whenever the relative change in the parameters

is small. We approximate the probability $Q_1(\mathbf{x})$ using the posterior distributions $\Delta(\delta)$ and $g(r)$ as follows:

$$Q_1(\mathbf{x}) \approx \sum_{\ell=-L}^L \sum_{\delta=-d}^d \prod_{i \in I_\delta} \varphi\left(\frac{x_i - y_i + r_\ell}{\sigma}\right) \prod_{i \notin I_\delta} \varphi\left(\frac{x_i - y_i}{\sigma}\right) \Delta(\delta) g(r_\ell), \quad (32)$$

where r_ℓ is ℓ standard deviations $\eta^{-1/2}$ above the mean ν . The posterior distribution in the case where either the meter readings are perfectly synchronized or the load-shed magnitude is deterministic can be computed as a special case of this procedure.

D. Problem parameters and available information

We consider the ADR maintenance problem in the following four settings:

- Case (a) Clocks synchronized and load-shed deterministic.
- Case (b) Clocks possibly mismatched with $d = 2$, but load-shed deterministic.
- Case (c) Clocks synchronized, but load-shed distributed $\mathcal{N}(\nu_0, \frac{1}{\eta_0})$.
- Case (d) Clocks possibly mismatched with $d = 2$, and load-shed distributed $\mathcal{N}(\nu_0, \frac{1}{\eta_0})$.

The parameter values for the numerical experiments were set as follows:

- (i) Observations in an hour-long DR event $m = 10$
- (ii) Probability of ADR failure $p = 0.05$
- (iii) Expected DR savings for utility $\lambda = 1$
- (iv) Cost of repair $c = 3\lambda = 3$
- (v) Discount factor $\beta = 0.9$
- (vi) Load-shed signal-to-noise ratio $\text{SNR} = 20 \log_{10}(\nu_0/\sigma) = \{-5, 0, 5\}$ dB, i.e. standard deviation σ of the meter noise $\approx \{1.78, 1, 0.56\}$ times the mean load-shed ν_0
- (vii) Load-shed standard deviation $\eta_0^{-1/2} = 0.1\sigma$

E. Discretization size n and sample size N

In Table I, we report the optimal threshold b^* , the approximate optimal value function $V(b)$ at belief state $b = 1$, and the elapsed time in seconds to approximate the expectation $\mathbb{E}_{\mathbf{x}}[V(\hat{\Gamma}_{\alpha_0 \mathbf{x}}(k))]$ and solve LP (18), as a function of the number n of points used to discretize the interval $[0, 1]$ and the number of samples N . From the results, it is clear that the number of samples N does not have a significant impact on the value function $V(1)$. We interrupted the computation of the optimal threshold and the optimal value for Case (d) for the largest-sized approximations, i.e. $N = 500\text{K}$. The solution time was too large to be of practical use. In Table I, we report this situation with an horizontal line. For the rest of the experiments in this section, we used $n = 100$ and $N = 5\text{K}$.

n	N	Case (a)	Case (b)	Case (c)	Case (d)	
100	5K	0.160	0.150	0.170	0.150	b^*
100	500K	0.160	0.150	0.160	—	
1K	5K	0.168	0.154	0.171	0.159	
1K	500K	0.163	0.154	0.166	—	
100	5K	8.374	8.558	8.498	8.713	$V(1)$
100	500K	8.309	8.550	8.448	—	
1K	5K	8.292	8.477	8.423	8.648	
1K	500K	8.275	8.515	8.412	—	
100	5K	1	2	8	125	time(s)
100	500K	25	132	763	—	
1K	5K	86	102	159	1291	
1K	500K	334	1373	8205	—	

TABLE I: Optimal threshold b^* , optimal value function $V(b)$ at belief state $b = 1$, and time in seconds to approximate $\mathbb{E}_{\mathbf{x}}[V(\hat{\Gamma}_{\alpha_0 \mathbf{x}}(k))]$ and solve LP (18), as a function of n and N .

F. Comparison with periodic review policies

In this section we report the performance of the proposed threshold policy with respect to the current practice of periodically inspecting the ADRs. Suppose the periodic review interval is q DR events. Then, the associated value function $U(q)$ satisfies the recursion

$$U(q) = -c + \lambda \sum_{j=0}^{q-1} \beta^j (1-p)^j + \beta^q U(q), \quad (33)$$

i.e. $U(q) = \frac{1}{1-\beta^q} \left(\frac{\lambda(1-\beta^q(1-p)^q)}{1-\beta(1-p)} - c \right)$. The value function $U(q)$ is maximized at $q^* = 18$, where $U(q^*) = 4.10$. In contrast, the POMDP value function is at least $V(0) = \{5.14, 5.37, 5.51\}$ for $\text{SNR} = \{-5, 0, 5\}$ dB, i.e. a relative improvement over $U(q^*)$ of at least $\{25, 31, 34\}\%$. These results clearly show that the POMDP-based method significantly outperforms any regular maintenance schedule.

G. ADR repair scheduling problem

In this section we report the numerical results for an ADR repair scheduling problem with $D = 100$ ADRs and $M = 5$ repair crews. We considered two variants of the ADR repair scheduling problem: one where the repair costs for all the ADRs were identical and set equal to 3λ , and another where repair cost for each ADR were sampled uniformly from the interval $(0, 6.5]\lambda$. For $c > 6.5\lambda$, the utility is better off not repairing the ADR at all.

When the states of the ADRs are fully observable and the repair costs are identical, the associated multi-dimensional DP can be reduced to a one-dimensional DP with state given by the number of non-working devices. Thus, one can easily compute the optimal policy using the value iteration algorithm. In this case, we compare the performance of the Whittle-index policy for the POMDP with this optimal policy. When the states of the ADRs are fully observable but the repair costs are not identical, it is not possible to compute the optimal policy. In this setting, we compare the Whittle-index policy for the POMDP with the Whittle-index policy for the full information MDP. Given that the partially observable Whittle-index policy can infer the state of an ADR only *after* observing the meter

readings during a DR event, a fairer comparison would be against the optimal policy of the MDP where the state of the ADR is observable only after a DR event. We call this problem, the *slow information MDP*.

We computed the value function of the Whittle-index policies using simulation. We simulated the performance of the policies over a time horizon of $T = 44$ DR events with a discount factor $\beta = 0.9$. This implies that the simulated T -horizon value function is within 1% of the infinite horizon value function. The results are averaged over 100 runs. The column marked “Whittle” in Table II reports the relative error of the full information Whittle-index policy with respect to the optimal full and slow information policy when all the repair costs are identical. The next four columns report the performance of the partial information Whittle-index policy for the four different cases listed in Section IV-D. Since the partial information policy does not have access to the state, its performance depends on the load-shed SNR. The row marked by SNR $[-5, 5]$ reports the performance of the policy when the SNR of each ADR is sampled uniformly from the interval $[-5, 5]$; all other rows report the performance when the SNR for all ADRs was set equal to the value corresponding to that row.

The full information Whittle-index policy is close to the optimal policy both in the full information and the slow information case. For very low noise levels, i.e. SNR=5dB, the performance of partial information Whittle-index policy is no more than 4.63% (resp. 1.57%) worse than that of the full (resp. slow) information optimal policy. On the other hand, for very high noise levels, i.e. SNR=-5dB, the partial information Whittle-index policy could be as bad as 10.75% (resp. 7.89%) suboptimal with respect to the full (resp. slow) information optimal policy. For reasonable noise levels, i.e. SNR=0dB for all ADRs, or randomly drawn from $[-5, 5]$ dB, the performance of the partial information Whittle index policy is within 6.94% (resp. 3.95%) of the full (resp. slow) information optimal policy. The results in this table suggest that the four cases in decreasing order of the sub-optimality of the Whittle policy are: (d), (b), (c), (a). Thus, it appears that the uncertainty in clock mismatch leads to an increased loss in performance as compared to the uncertainty in the load-shed.

Table III shows the relative error of the partial information Whittle-index policy with respect to the full and slow information Whittle-index policy when the repair costs are not identical. Assuming that the full information Whittle-index policy is close to the optimal policy also in this case, the results in this table are similar to those observed in the case of identical repair costs.

V. CONCLUSIONS

In this paper, we formulated and solved the ADR repair scheduling problem where the goal is to maintain D ADRs using at most $M (\ll D)$ maintenance crews and the ADR state is only partially observable via noisy meter readings. We formulated this problem as a restless bandit problem. We showed that the ADR repair is Whittle-indexable, and therefore, one can very efficiently compute a good heuristic policy by suitably

SNR	err (%)					
	Whittle	Case (a)	Case (b)	Case (c)	Case (d)	
5	1.28	4.63	4.62	4.60	4.61	Full
0	1.28	5.38	5.92	5.77	6.56	
-5	1.28	8.48	9.63	9.23	10.75	
[-5,5]	1.28	5.76	6.3	6.14	6.94	
5	1.66	1.57	1.56	1.54	1.55	Slow
0	1.66	2.35	2.9	2.75	3.57	
-5	1.66	5.55	6.73	6.32	7.89	
[-5,5]	1.66	2.74	3.3	3.13	3.95	

TABLE II: Identical repair costs. Average error of the full, slow, and partial information Whittle-index policies w.r.t. the optimal full and slow information policies. See Section IV-D for the details on partial information Cases (a)-(d).

SNR	err (%)				
	Case (a)	Case (b)	Case (c)	Case (d)	
5	3.28	3.32	3.32	3.39	Full
0	4.26	4.87	4.7	5.45	
-5	7.15	8.04	7.73	8.93	
[-5,5]	4.61	5.15	4.96	5.66	
5	-0.15	-0.11	-0.1	-0.03	Slow
0	0.87	1.5	1.32	2.1	
-5	3.86	4.78	4.46	5.7	
[-5,5]	1.23	1.78	1.59	2.32	

TABLE III: Non-identical repair costs. Average error of the partial information Whittle-index policy with respect to the full and slow information Whittle-index policies. See Section IV-D for the details on partial information Cases (a)-(d).

decomposing the D -ADR repair scheduling problem into D single-ADR maintenance problems. We showed that the optimal solution of the single-ADR maintenance problem is a threshold policy, where it is optimal to send the maintenance crew as soon the belief state drops below a threshold. Using the structure of the single-ADR optimal policy we showed that the Whittle index as a function of the belief state of an ADR can be computed via a single binary search. We explored the performance of the Whittle-index policies when the meter and utility clocks are (resp. not) synchronized and the load-shed is random (resp. deterministic) (see Section IV-D for the details of the four cases). We also developed a hierarchical Bayesian method for computing the joint posterior distribution of the load-shed and clock mismatch. Our numerical experiments suggest that, when the level of noise in the meter readings is on average of the same size of the load shed, the Whittle-index policy is at most 3.95% suboptimal. This problem and its solution was motivated by our research collaboration with AutoGrid, a software provider for managing DR programs.

REFERENCES

- [1] Federal Energy Regulatory Commission. National Assessment & Action Plan on Demand Response.
- [2] J Wells and D Haas. *Electricity markets: consumers could benefit from demand programs, but challenges remain*. DIANE Publishing, 2004.
- [3] Mary Ann Piette, Girish Ghatikar, Sila Kiliccote, David Watson, Ed Koch, and Dan Hennage. Design and operation of an open, interoperable automated demand response infrastructure for commercial buildings. *Journal of Computing and Information Science in Engineering*, 9(2), 2009.

- [4] Open ADR alliance: Overview. <http://www.openadr.org/overview>.
- [5] Office of Electricity Delivery & Energy Reliability, U.S. Department of Energy. SGIG accelerates grid modernization in Minnesota. Case study: Minnesota Power, 2012.
- [6] Silver Spring Networks. Breathing new life into legacy demand response (white paper), 2013.
- [7] Silver Spring Networks. The business case for DLC replacement (white paper), 2013.
- [8] CH Papadimitriou and JN Tsitsiklis. The complexity of Markov decision processes. *Mathematics of Operations Research*, 12(3):441–450, 1987.
- [9] O Madani, S Hanks, and A Condon. On the undecidability of probabilistic planning and infinite-horizon partially observable Markov decision problems. In *AAAI/IAAI*, pages 541–548, 1999.
- [10] J Endrenyi, S Aboresheid, RN Allan, GJ Anders, S Asgarpour, R Billinton, N Chowdhury, EN Dyalynas, M Fipper, RH Fletcher, et al. The present status of maintenance strategies and the impact of maintenance on reliability. *Power Systems, IEEE Transactions on*, 16(4):638–646, 2001.
- [11] P Whittle. Restless bandits: Activity allocation in a changing world. *Journal of Applied Probability*, 1988.
- [12] JC Gittins. Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1979.
- [13] CH Papadimitriou and JN Tsitsiklis. The complexity of optimal queueing network control. *Mathematics of Operations Research*, 24(2):293–305, 1999.
- [14] KD Glazebrook, D Ruiz-Hernandez, and C Kirkbride. Some indexable families of restless bandit problems. *Advances in Applied Probability*, 2006.
- [15] D Braziunas. POMDP solution methods. *University of Toronto, Tech. Rep*, 2003.
- [16] KJ Åström. Optimal control of Markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications*, 10(1):174–205, 1965.
- [17] George E Monahan. State of the art: A survey of partially observable Markov decision processes: Theory, models, and algorithms. *Management Science*, 28(1):1–16, 1982.
- [18] EJ Sondik. The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs. *Operations Research*, 26(2):282–304, 1978.
- [19] S Boyd and L Vandenberghe. *Convex optimization*, chapter 3.2. Cambridge University Press, 2009.
- [20] ML Puterman. *Markov decision processes: discrete stochastic dynamic programming*, volume 414, chapter 6. John Wiley & Sons, 2009.
- [21] IM Sobol. On quasi-Monte Carlo integrations. *Mathematics and Computers in Simulation*, 47(2):103–112, 1998.
- [22] L Devroye. Sample-based non-uniform random variate generation. In *Proceedings of the 18th conference on Winter simulation*, pages 260–265. ACM, 1986.
- [23] CM Bishop. *Pattern recognition and machine learning*, volume 1, pages 152–160, 461–516. Springer New York, 2006.

Carlos Abad is a PhD candidate in the Industrial Engineering and Operation Research Department at Columbia University. His research interests include convex optimization, compressive sensing, Bayesian inference, decision making under uncertainty, and their applications to the operation and economics of electric grids.

Garud Iyengar received a Ph.D. in Electrical Engineering from Stanford University. He joined Columbia University’s Industrial Engineering and Operations Research Department in 1998 and teaches courses in asset allocation, asset pricing, simulation and optimization. His research interests include convex, robust and combinatorial optimization, queueing networks, mathematical and computational finance, communication and information theory.